Nucleic Acids Research, Vol. 19, No. 24 6863–6869

Purification and substrate specificity of a T4 phage intron-encoded endonuclease

Frederick K.Chu, Frank Maley, Ai-Min Wang, Joan Pedersen-Lane and Gladys Maley
Wadsworth Center for Laboratories and Research, New York State Department of Health, Albany, NY 12201-0509, USA

Received August 28, 1991; Revised and Accepted November 19, 1991

ABSTRACT

The T4 phage *td* intron-encoded endonuclease (I-Tev I) cleaves the intron-deleted *td* gene (*td*ΔI) 23 nucleotides upstream of the intron insertion site on the noncoding strand and 25 nucleotides upstream of this site on the coding strand, to generate a 2-base hydroxyl overhang in the 3′ end of each DNA strand. I-Tev I-157, a truncated form in which slightly more than one third (88 residues) of the endonuclease is deleted, was purified to homogeneity and shown to possess endonuclease activity similar to that of I-Tev I, the full-length enzyme (245 residues). The minimal length of the *td*ΔI gene that was cleaved by I-Tev I and I-Tev I-157 has been determined to be exactly 39 basepairs, from −27 (upstream in exon1) to +12 (downstream in exon2) relative to the intron insertion site. Similar to the full-length endonuclease, I-Tev I-157 cuts the intronless thymidylate synthase genes from such diverse organisms as *Escherichia coli*, *Lactobacillus casei* and the human. The position and nature of the *in vitro* endonucleolytic cut in these genes are homologous to those in *td*ΔI. Point mutational analysis of the *td*ΔI substrate based on the deduced consensus nucleotide sequence has revealed a very low degree of specificity on either side of the cleavage site, for both the full-length and truncated I-Tev I.

INTRODUCTION

An endonuclease, encoded by the intron open reading frame of 245 codons in the T4 bacteriophage thymidylate synthase gene (*td*) (1,2), cleaves the intron-deleted version of the *td* gene (*td*ΔI) (3) at a unique site centered at 24 bp upstream of the exon splice junction, the site in which the 1016-bp intron resides when present (4,5). The endonucleolytic cut occurs at 23 residues upstream of this junction on the noncoding strand, and 25 residues on the coding strand, forming a 2-base stagger with 3′ hydroxyl overhangs (4). The biological role of I-Tev I is unclear, but genetic experiments have shown that I-Tev I is essential in the nonreciprocal transfer of the *td* intron to the *td*ΔI gene (6). The intron transfer is initiated by an endonucleolytic event believed to be catalyzed by I-Tev I, in the vicinity of the intron insertion site, and followed by duplicative recombination which results in co-conversion of flanking exon sequences (7,8).

We have described in our earlier work (3) a truncated form of I-Tev I which is a fusion protein containing in its amino end an additional 17 amino acid residues of T7 gene 10 origin intrinsic in the expression plasmid pET3c. Since the fusion protein contains the first 157 residues of I-Tev I as shown in the present work, it will be referred to as I-Tev I-157 to distinguish it from the full-length endonuclease I-Tev I. We have purified I-Tev I-157 to homogeneity and have shown that it possesses identical substrate specificity to that of I-Tev I. We have previously estimated an upper limit of 87 basepairs for the *td*ΔI recognition sequence for cleavage by either form of the intron endonuclease (4). In this work using *in vitro* constructed heteroduplex and homoduplex substrates, we have accurately determined the minimal *td*ΔI substrate length to be 39 basepairs. In addition, both forms of the endonuclease can cleave other intronless thymidylate synthase genes. With the deduced consensus nucleotide sequence in these substrates as a model for single base mutational analysis, we have uncovered a low degree of specificity in the nucleotides around the cleavage site.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, Enzymes and Chemicals

Escherichia coli TG1, obtained from Amersham, was used for propagating M13mp8 recombinant phages. Host strain HMS174 with and without the pLysS plasmid (kindly provided by F.W. Studier, Brookhaven National Laboratory, Upton, NY) (9,10). was used in the biosynthesis of I-Tev I and I-Tev I-157 endonucleases from the plasmids pETdIrf and pETdIrf-157 (labeled pETdIrf in ref. 3), respectively. The pET3c plasmid and CE6λ phage were obtained also from F.W. Studier. The following recombinant plasmids were used as substrates for the endonucleases: pUC*td*ΔI containing the intron-deleted thymidylate synthase (TS) gene from T4 phage (11); pBSTAH containing the intronless gene from *E. coli* (12); pKPTS containing the intronless gene from *Lactobacillus casei* (13) and pWHTS (14) containing the cDNA encoding human TS (15) were both obtained from D.V. Santi, U. Cal., San Francisco; and pBSthyP3 (prepared in our laboratory) containing the intronless TS gene from the *Bacillus subtilus* phage ϕ3T (16). Restriction enzymes and DNA modifying enzymes were purchased from several suppliers. [γ-^{32}P]ATP (>5000 Ci/mmol; 1 Ci = 37 GBq) and deoxyadenosine [α-[^{35}S]thio]triphosphate (>4000

Ci/mmol) were from Amersham. Oligodeoxyribonucleotides were synthesized with a model 381A DNA synthesizer from Applied Biosystems.

Endonuclease Purification

The I-Tev I-157 endonuclease was purified from *E. coli* HMS174 cells harboring the pETdIrf-157 plasmid following induction by infection with CE6λ phage as described previously (3). The full-length endonuclease, I-Tev I, was induced from pETdIrf using the same procedure. The cells were harvested from a 1-liter culture by centrifugation at 8000×g for 15 min and the cell pellet was resuspended in 20 ml of 10 mM Tris.HCl, pH 7.5, 10 mM EDTA and centrifuged at 8000×g for 15 min. At this stage, the cells could be stored at -20°C without significant loss of endonuclease activity. The washed cell pellet (~2 g) was taken up in 20 ml of lysis buffer containing 50 mM potassium phosphate, pH 7.5, 5% triton X-100, 8% sucrose, 45 mM EDTA, pH 8.0, and 2.6 mg of lysozyme. The resulting suspension was incubated at 4°C for 16 h and then sonicated 5 times for 1 min each using the micro-tip of a Vibra-cell VC 600 processor (Sonics and Materials, Danbury, CT) and centrifuged at 27000×g for 20 min. The inclusion body-containing pellet was resuspended in 10 ml of lysis buffer and centrifuged again at 27000×g for 15 min. The resulting pellet was washed with a solution of 10 mM Tris.HCl, pH 7.5, 10 mM EDTA, centrifuged and the pellet solubilized in 4 ml of a solution containing 6 M guanidine.HCl, 50 mM potassium phosphate, pH 7.5, and 100 mM 2-mercaptoethanol. After 2 h at room temperature, the mixture was centrifuged at 10000×g for 15 min. The clarified supernatant was applied to a column (1.5×85 cm) of Biogel A-0.5 M pre-equilibrated in a buffered solution of 4 M guanidine.HCl, 50 mM potassium phosphate, pH 7.5, 100 mM 2-mercaptoethanol and eluted in the same buffer at a flow rate of 12 ml per h. Two-ml fractions were collected and assayed for absorbance at 280 nm, as well as for cleavage of the *td*ΔI DNA substrate after reactivation of the enzyme by dilution or dialysis. The enzyme fractions were pooled and dialyzed overnight at 4°C against two 2-liter changes of a buffer containing 25 mM potassium phosphate, pH 7.5, 20 mM 2-mercaptoethanol and 500 mM NaCl. In some cases, the dialysate contained a white precipitate which was removed by centrifugation. The dialysate (20 ml) was concentrated about 10-fold using an Amicon stirred cell with a PM10 membrane. Solid ammonium sulfate was added to this enzyme solution to a final concentration of 1 M which was loaded onto a phenyl-sepharose CL-4B column (1×3 cm) pre-equilibrated with a buffer containing 50 mM potassium phosphate, pH 7.5, 1 M ammonium sulfate and 20 mM 2-mercaptoethanol. The column was washed with the same buffer until the A_{280} decreased to background and 2-ml fractions were collected. The column was developed with buffer containing 0.8 M ammonium sulfate, then 0.5 M ammonium sulfate and finally buffer alone. The endonuclease activity was present in the 0.5 M fractions, which were pooled (4 to 6 ml) and concentrated to 1 ml in a centricon-10. The concentrate was mixed with an equal volume of glycerol and stored at −20°C.

In vitro Endonuclease Assay with Plasmid DNA Substrates

Plasmid DNA to be tested as a substrate for I-Tev I or I-Tev I-157 endonuclease was linearized with *Eco*RI, which cleaves pUC*td*ΔI in the multiple cloning site. The DNA was purified using the Geneclean procedure (BIO 101 Inc., La Jolla, CA). The assay mixtures (15 μl) contained 50 mM Tris.HCl, pH 7.5, 10 mM $MgCl_2$, 50 mM NaCl, 5 mM dithiothreitol, 0.1 to 0.15 pmol DNA substrate, and up to 1.9 μg of I-Tev I or I-Tev I-157. The time and temperature of incubation varied, but for routine assays the reactions were usually for 10 min at 37°C, at which time they were stopped by the addition of 5 μl of stop-load buffer containing 50 mM EDTA, 5% SDS, 25% glycerol and 0.1% bromophenol blue. The extent of endonucleolytic reaction was analyzed on 1% agarose gels in TBE buffer (0.1 M Trizma base, 0.1 M boric acid, 2 mM Na_2 EDTA). DNA bands were visualized by staining with ethidium bromide and photographed using a Polaroid MP4 apparatus. One unit of I-Tev I endonuclease is defined as the amount of enzyme required to cleave 1 pmole of the linear pUC*td*ΔI DNA substrate per h at 37°C at pH 7.5. Protein concentrations were determined by a modification of the method of Beardon (17) using bovine serum albumin as a standard.

Restriction Analysis of ^{35}S-labeled cDNA Substrates with Endonuclease

Labeled cDNA products were prepared in the presence of deoxyadenosine [α-[^{35}S]thio]triphosphate with and without the individual dideoxyribonucleoside triphosphates (A,C,G,T) using Sequenase (United States Biochemical Corp.) and the appropriate oligonucleotide primer to copy from either single-stranded (ss) or double-stranded (ds) DNA templates (4). The cDNA reactions without dideoxyribonucleotide chain-terminators were used as substrates for I-Tev I or I-Tev I-157 endonuclease, and those with chain-terminators were used as sequence ladders to pinpoint the endonucleolytic cut. The processing and treatment of cDNAs with the endonuclease was basically as described previously (4) except that the cleavage was carried out at 37°C for 10 min with 0.1 μg of endonuclease. The products were analyzed on 8% polyacrylamide/6 M urea sequencing gel (32×42 cm) in TBE buffer.

Restriction Analysis of ^{32}P-labeled Annealed Complexes of Oligodeoxyribonucleotide and M13td ΔI ssDNA with Endonuclease

One-tenth picomole of ^{32}P 5′ end-labeled oligodeoxyribonucleotide was incubated at 75°C for 3 min with 0.3 pmol of M13*td*ΔI ssDNA in annealing buffer (40 mM Tris.HCl, pH 9, 50 mM NaCl) in a final volume of 5 μl. The annealed mixture was cooled slowly to room temperature (approx. 30 min) and divided into two 2.5-μl aliquots. To one aliquot were added 1.25 μl of buffer containing 40 mM Tris.HCl, pH 9, 50 mM NaCl, 30 mM $MgCl_2$ and 30 mM dithiothreitol, and to the other, the same buffer with 0.15 μg of I-Tev I or 0.1 μg of I-Tev I-157. After incubating for 10 min at 37°C, 2.25 μl of stop solution (95% formamide, 20 mM Na_2EDTA, 0.05% bromophenol blue, 0.05% xylene cyanol) were added, followed by heating for 3 min at 80°C. The samples were chilled briefly on ice and 3-μl aliquots were subjected to electrophoretic analysis on 15% polyacrylamide/6 M urea gel (16×12 cm) in TBE buffer.

Restriction Analysis of Mutant Substrate Sequences with Endonuclease

Single-base substitutions at several nucleotide positions on either side of the I-Tev I endonucleolytic cleavage site in the *td*ΔI gene were constructed using M13*td*ΔI ssDNA as the template. The protocol for the oligonucleotide-directed mutagenesis (18,19) as

outlined by Amersham was followed. Single-stranded DNAs were prepared from the M13*td*ΔI mutant constructs, their mutant sequences confirmed by dideoxy sequencing, and then used as templates for the synthesis of ^{35}S-labeled cDNAs primed with an exon2 oligonucleotide (16 mer: 5'-ACACATCTTAGCTA-CA-3') starting at 57 bases downstream of the intron insertion site. The resulting dsDNA products were tested as substrates for intron endonuclease as described (ref. 4 and this work).

Other Methods

Transfer of DNA from agarose gel to Hybond-N membrane (Amersham) by capillary blotting and subsequent hybridization analysis of the transferred material with ^{32}P-labeled oligonucleotides were performed as described previously (4). The oligonucleotides used in annealing to M13*td*ΔI ssDNA and as probes in hybridization analysis were phosphorylated at the 5' end with T4 polynucleotide kinase in the presence of [γ-^{32}P]ATP. Gel electrophoretic analysis of protein samples was performed in 15% polyacrylamide slab gels employing the discontinuous system of Laemmli (20).

RESULTS

Purification of I-Tev I-157, a Truncated Form of I-Tev I Endonuclease

We have previously described a form of I-Tev I endonuclease that appeared to be truncated at its carboxyl end (3). We have since determined that this was not a case of proteolytic processing but was due to a C to T mutation in the first nucleotide of the 158th codon, thus converting CAA (glutamine) to TAA (termination) during the course of cloning the IRF gene from M13 into pET3c. Expression from the mutant IRF yielded I-Tev I-157 in which the C-terminal 88 amino acids were missing, resulting in a protein containing 157 residues instead of the entire 245 residues of the I-Tev I endonuclease. Surprisingly, induced extracts containing this significantly truncated form of I-Tev I showed comparable *td*ΔI cleavage activity to that of the full-length 245-residue long enzyme. Furthermore, I-Tev I-157 in crude or partially purified fractions exhibited much higher stability than the full-length endonuclease, which prompted us to purify this more stable form of I-Tev I. The same purification procedure (see Materials and Methods) was used for the preparation of the full-length and the truncated forms of the endonuclease.

Following induction of the pETdIrf-157 plasmid in *E. coli* HMS174 cells by infection with CE6λ phage, about 10% of the total cellular protein was I-Tev I-157 (Fig. 1, lane 1), which was present almost entirely in inclusion bodies. Isolation of the inclusion bodies facilitated a 7 to 8 fold purification of the endonuclease (lane 3), which unfortunately contained significant exonuclease activity. Chromatography on Biogel A-0.5 M and then on phenyl-sepharose removed this undesired activity and yielded at least 98% pure I-Tev I-157 endonuclease (lane 4). At this stage, I-Tev I-157 showed 50% higher specific activity (2800 units/mg protein) than full-length I-Tev I due to its one-third smaller size, and cleaved the *td*ΔI substrate in a manner identical to that of the full-length endonuclease. Substrate specificity studies were carried out with both I-Tev I and I-Tev I-157. As similar results were obtained with the two forms, both in terms of the cleavage site and cleaved structure in the substrate molecules, only data with I-Tev I-157 are presented in this work.

The fact that the enzyme protein is purified from an inclusion body complex introduces both positive and negative aspects to the isolation. In the former case, many undesirable proteins are removed providing a highly enriched but inactive endonuclease fraction. In the latter case, reactivation of the enzyme on removal of the solubilizing agent (guanidine.HCl) and purification to homogeneity may not remove residual inactive enzyme, making it impossible to determine its true specific activity. For this reason, and the possibility that during the renaturation process, some or all of the enzyme molecules might be active to varying degrees due to less than perfect refolding, our calculated specific activity would be an underestimate. However, since the assay does not provide, at present, a precise quantitative measure of the rate of cleavage, this becomes a moot point until such time as the assay is improved.

Determination of Minimal tdΔI Gene Length for Cleavage by I-Tev I-157 Endonuclease

It was shown earlier by us that a relatively long 87-bp stretch in the *td*ΔI gene contains the recognition sequence for the *td* intron endonuclease (4). To determine the minimal length of this gene

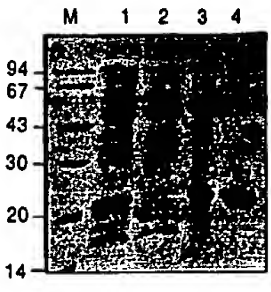

Figure 1. Polyacrylamide gel electrophoretic analysis of the I-Tev I-157 protein fractions at various stages of purification. Size of the molecular weight markers (Pharmacia) are indicated on the left of the 'M' lane which contains phosphorylase b (94 kDa), bovine serum albumin (67 kDa), ovalbumin (43 kDa), carbonic anhydrase (30 kDa), soybean trypsin inhibitor (20 kDa), and α-lactabumin (14 kDa). Lane 1, sonicated extract; lane 2, supernatant after centrifugation; lane 3, guanidine.HCl solubilized pellet; lane 4, phenyl-sepharose purified fraction (0.5 M ammonium sulfate eluate). Electrophoresis was carried out in 15% polyacrylamide-SDS gel according to Laemmli (20).

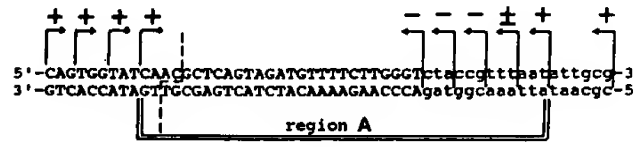

Figure 2. Restriction analysis of tdΔI cDNA substrates. Double-stranded pUC*td*ΔI plasmid DNA was used as template for cDNA synthesis in the presence of deoxyadenosine [α-[^{35}S]thio]triphosphate according to the Sequenase protocol of United States Biochemical Corp. The ^{35}S-labeled cDNA was treated with I-Tev I and with I-Tev I-157 endonucleases and the resulting products were analyzed on an 8% sequencing gel. The sequence ladder, consisting of ^{35}S-labeled cDNA synthesized from pUC*td*ΔI in the presence of the individual dideoxynucleotide (A,C,G,T) chain terminators, was used to locate the cleavage site. The degrees of cDNA cleavage by the *td* intron endonucleases (+, ≥70%; ±, ≤10%; −, 0%) were estimated from an autoradiogram of the gel. The direction of priming and the 5' end of the primers are indicated by the arrows. The dashed line shows the site of endonucleolytic cleavage by the endonucleases. Region A is 39 basepairs in length and spans 27 basepairs of exon1 (upper case letters) and 12 basepairs of exon2 (lower case letters). The exon boundary is as defined in the intron-containing *td* gene (1).

that can be cleaved by the endonuclease, we constructed heteroduplex DNA molecules containing progressively shortened double-stranded regions in the 87-bp stretch. These were then tested as substrates for the purified endonucleases. We achieved this in two phases. In the first phase, we synthesized oligodeoxyribonucleotides (11 to 15 bases in length) for priming pUC*td*ΔI dsDNA in both directions at decreasing distances from the endonucleolytic cleavage site. The primed sites were separated from each other by 3 or 6 bases and the resulting cDNAs produced in the presence of deoxyadenosine [α-[^{35}S]thio]triphosphate were treated with I-Tev I or I-Tev I-157. Their degree of cleavage (%) was estimated relative to cDNAs primed with an 18 mer oligonucleotide complementary to exon1 at 73 residues upstream, and a 16 mer complement of exon2 at 57 residues downstream, respectively, of the cleavage site. Figure 2 summarizes the results of the study. It was surprising to observe that the upstream sequence in the coding strand could be trimmed to 2 residues from the staggered cut without appreciable influence on *in vitro* cleavage efficiency. On the other hand, the downstream sequence required for efficient cleavage included the first 12 residues of exon2 (37 residues downstream of the cleavage site on the coding strand). Trimming to 9 residues in exon2 (34 residues downstream of the cleavage site) reduced the efficiency by almost ten-fold. These results show that a 39-bp stretch of deoxynucleotides (region A in Fig. 2) in the *td*ΔI gene contains the necessary substrate information for recognition and cleavage by the intron endonuclease.

To determine if this 39-bp region is the absolute substrate length, the second phase of experimentation was employed. Oligonucleotides within region A were synthesized, labeled with ^{32}P on their 5′ ends and then annealed to M13*td*ΔI ssDNA. The resulting heteroduplexes were tested directly as substrates for both I-Tev I-157 and the full-length endonuclease. The results for three oligonucleotides of interest are shown in Figure 3. Panel A characterizes the oligos: I, a 39 mer representing region A; II, a 38 mer without the upstream (5′) ultimate G residue; and III, a 38 mer without the downstream (3′) ultimate A residue. As shown in panel B, the heteroduplex containing the 39 mer oligo I (lane 1) was readily cleaved by either endonuclease (data shown for I-Tev I-157) converting the oligo to a 37 mer (lane 2), consistent with hydrolysis of the phosphodiester bond between the second (T) and third (T) nucleotides in oligo I. By contrast, both 38 mer oligos II (lane 3) and III (lane 5) which were shorter than oligo I by one residue in the 5′ and 3′ ends, respectively, were not cleaved (lanes 4 and 6, respectively). The requirement for double-stranded DNA structure was reaffirmed by the fact that in the absence of M13*td*ΔI ssDNA, oligo I was not cleaved (panel C).

Since the heteroduplex substrate formed from oligo I contained extra *td*ΔI sequences in the M13 recombinant strand outside of region A which may affect the recognition and cleavage by the intron endonuclease, we decided to test the 39-bp homoduplex

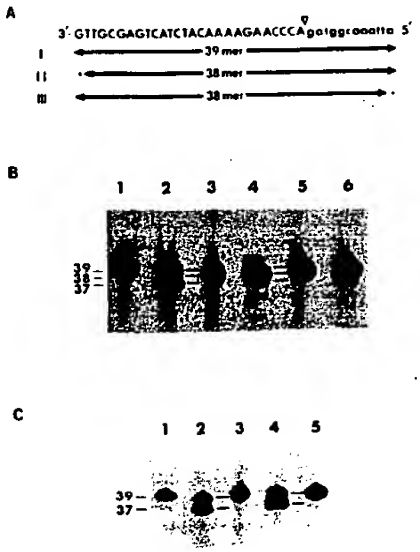

Figure 3. Restriction analysis of annealed DNA complexes as substrates for I-Tev I-157 endonuclease. (A) Synthetic oligodeoxyribonucleotides used in heteroduplex formation with M13*td*ΔI ssDNA. The intron insertion site (▽). exon1 (upper case letters) and exon 2 (lower case letters) sequences are indicated as they appear in the *td* gene. (B) Gel profile of *td*ΔI heteroduplexes before and after treatment with I-Tev I-157 endonuclease. Oligonucleotides I, II and III were first ^{32}P-labeled on their 5′ ends, annealed to M13*td*ΔI ssDNA, and then treated with the endonuclease. Substrates (lanes 1,3,5) and I-Tev I-157-treated products (lanes 2,4,6) were resolved on a 15% polyacrylamide-urea slab gel and subjected to autoradiography. The lanes contain: 39 mer oligo I (lanes 1,2); 38 mer oligo II (lanes 3,4); and 38 mer oligo III (lanes 5,6). The position of 37, 38 and 39 mer oligonucleotides in the gel are indicated. (C) I-Tev I-157 endonuclease cleavage of both *td*ΔI heteroduplex and homoduplex substrates. The heteroduplex (lanes 1,2) contained ^{32}P-labeled 39 mer oligo I and M13*td*ΔI ssDNA; and the homoduplex (lanes 3,4), ^{32}P-labeled 39 mer oligo I and its 39 mer unlabeled complement. Substrates (lanes 1,3) and endonuclease-treated products (lanes 2,4,5) were analyzed as in (B). Lane 5 is ^{32}P-labeled oligo I alone incubated with I-Tev I-157 endonuclease.

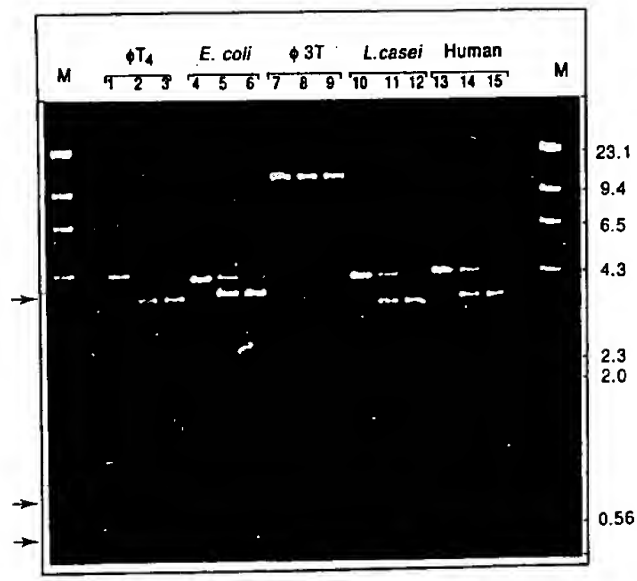

Figure 4. Cleavage of linearized plasmid substrates derived from various thymidylate synthase genes by I-Tev I-157 endonuclease. The various DNAs (0.1 pmol) were incubated in 50 mM Tris.HCl, pH 7.5, 10 mM $MgCl_2$, 5 mM dithiothreitol, 50 mM NaCl in the absence (lanes 1, 4, 7, 10, 13) and in the presence of 1.9 μg of I-Tev I-157 for 5 min (lanes 2, 5, 8, 11, 14) and 15 min (lanes 3, 6, 9, 12, 15) at 24°C. The constructions for the various TS DNAs are T4 phage, pUC*td*ΔI; *E. coli*, pBSTAH; *Bacillus* phage, pBSthyP3; *L. casei*, pKPTS; human, pWHTS. The first three constructs were linearized with *Eco*RI and the last two, with *Hin*dIII. The linear DNAs were purified with Geneclean (BIO 101). Substrates and endonuclease-treated products were analyzed on 1% agarose gel with DNA size markers in the leftmost and rightmost lanes.

as a substrate. An oligonucleotide complementary to oligo I was prepared, annealed to ^{32}P-labeled oligo I and then treated with I-Tev I or I-Tev I-157. As shown in panel C for I-Tev I-157, the 39-bp homoduplex was cleaved (lane 4) just as efficiently as the heteroduplex (lane 2) indicating that the additional sequence outside of region A in the heteroduplex substrate was not required and did not affect the recognition and cleavage by either endonuclease. These results demonstrate that the 39-bp region A in *td*ΔI contains all the necessary substrate information for recognition and cleavage *in vitro* by I-Tev I-157 or I-Tev I. In fact, region A is the shortest *td*ΔI substrate for the intron endonuclease since reducing it further in size inactivates it as a substrate.

I-Tev I-157, Like I-Tev I, Cleaves Other Thymidylate Synthase Genes

We have observed previously that I-Tev I-157 can cleave the *E. coli thyA* gene (3) which is less than 70% homologous (26 out of 39 bases) to the comparable region A in the *td*ΔI gene. To obtain information about the consensus substrate sequence for the endonuclease, we decided to survey other intronless or intron-deleted thymidylate synthase (TS) genes as possible substrates. Four TS genes were available and included those from *E. coli*, *Lactobacillus casei*, human cDNA and the *Bacillus* phage, ϕ3T. It was found that I-Tev I-157, as well as I-Tev I (data not shown), introduced a single cut in the synthase genes of *E. coli*, *Lactobacillus casei* and human cDNA, regardless of whether they were linear (Fig. 4) or circular substrates (data not shown). However, the intronless synthase gene of 3T was not a substrate, nor was the T4 phage *td* gene (3,4) where the 39-bp recognition region is interrupted by an 1016-bp intron (1) containing the IRF gene coding for I-Tev I (2,3).

To elucidate the structure of the cleavage site in each dsDNA substrate, cDNA was made from each strand primed with the appropriate oligonucleotides from both upstream and downstream regions in the presence of [α-^{35}S]dATP. The cDNA products were then treated with purified I-Tev I-157 and electrophoresed on an 8% sequencing gel next to the corresponding sequence ladder. Only I-Tev I-157 was used as endonuclease in this experiment because it appears to be more stable than the full-length I-Tev I and both forms show the same substrate specificity (Fig. 4). As shown in Figure 5, I-Tev I-157 cleaved three of the synthase gene substrates (*E. coli*, *L. casei*, human cDNA) in exactly the same manner as it did the *td*ΔI gene, and in all cases generated a 2-base staggered cut with 3′ hydroxyl overhangs (Fig. 5A to D) which could be ligated by the action of T4 phage DNA ligase (unpublished results). Furthermore, the endonucleolytic cut in each case was located at the same position in the homology map of the TS gene. As observed with recombinant plasmid DNA as substrate, the cDNA derived from ϕ3T gene was not cleaved (Fig. 5E).

Mutational Studies of the Cleavage Sequence

Based on the sequence of the described thymidylate synthase genes which were shown to be substrates of the intron endonuclease (Fig. 5), a consensus sequence of 12 nucleotides for the cleavage site was deduced (Table 1, II). To test the

Table 1. Mutational studies of cleavage site sequence in the *td*ΔI gene. Single-base substitutions were introduced in a 12-base sequence (noncoding strand sequence shown) centered at the cleavage site, using the M13 ss mutagenesis system (see Materials and Methods). The resulting M13 *td*ΔI mutant ssDNAs were used as templates for synthesis of ^{35}S-labeled cDNAs which were then tested for susceptibility to cleavage by I-Tev I-157: +, cleavage efficiency similar to that of the wild-type sequence; ±, cleavage efficiency less than that of the wild-type (≤ 50%). The arrows indicate where the endonucleolytic cut occurs in the noncoding DNA strand. The nucleotides in the consensus sequence are numbered −1 to −6 for upstream and +1 to +6 for downstream residues, starting from the center of the endonucleolytic cut. The endonuclease generates a 2-base stagger which is represented by positions −1 and +1.

I.	POSITION	6	5	4	3	2	-1	+1	2	3	4	5	6
II.	CONSENSUS	5′ · T · A · Py · C · A · Pu · C/A ↓ G · N · T · C · N · 3′											
III.	tdΔI	5′ · T · A · T · C · A · A · C ↓ G · C · T · C · A · 3′											
IV.	MUTATION A	+		+				+		+			
	C	+	+	+			+			+			
	G	+		+			+	+		±			
	T						+	+					

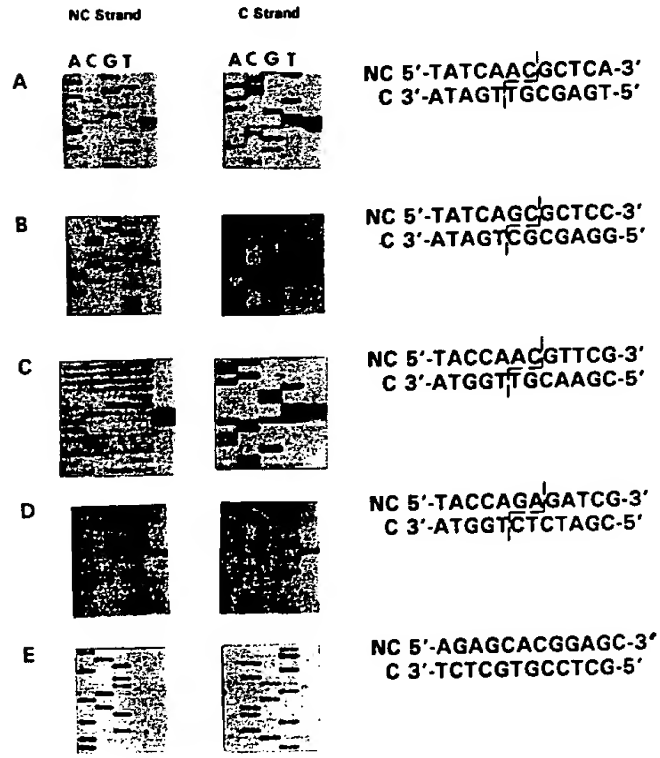

Figure 5. Survey of substrates for I-Tev I-157 endonuclease: structural analysis of the cleavage site in thymidylate synthase genes. The thymidylate synthase genes tested were: (A) T4 phage (pUC*td*ΔI); (B) *E. coli* (pBSTAH); (C) *L. casei* (pKPTS); (D) human cDNA (pWHTS); and (E) *Bacillus* phage ϕ3T (pBSthyP3). Synthesis of cDNA from each recombinant plasmid dsDNA and subsequent restriction with I-Tev I-157 endonuclease were as described in the legend to Figure 2. The synthetic oligonucleotides for priming each template from both directions were designed such that they started at 60 to 80 residues from the putative cleavage site anticipated from a comparison with the known position of cleavage in the *td*ΔI gene. Each set of 5 lanes shows the sequence ladder (A,C,G,T in lanes 1 to 4) and I-Tev I-157-treated cDNA (lane 5) derived from the corresponding thymidylate synthase gene. The sequences on the right side were read from the gel panels on the left side (cDNA bands are highlighted with dots). The symbols are: NC, noncoding strand; C, coding strand; —, cleavage site.

stringency of this sequence, the *td*ΔI gene was used as template for oligonucleotide-directed single-base substitution mutations in and around the cleavage site (Table 1, III). Sixteen *td*ΔI mutants were generated and subsequently tested as substrates for I-Tev I-157. The results are listed in Table 1, IV. The sixteen mutants are represented by nucleotide changes in six of the twelve positions in the sequence. Seven mutants located in the triplet of bases (positions −6 to −4), at one codon upstream of the cleavage site, exhibited no significant change in the efficiency of cleavage by the endonuclease when compared to the wild type substrate sequence. Surprisingly, six mutants located in the doublet of bases (positions −1 and +1) representing the stagger in the cleavage site also exhibited little change in cleavage efficiency. Only one of three mutants located at position +4, one codon downstream of the cleavage site, showed a significant reduction in cleavage efficiency: a T to G change at this position resulted in a 50% loss in cleavage efficiency.

DISCUSSION

The endonuclease, I-Tev I, is encoded by an intron reading frame (IRF) in the T4 phage thymidylate synthase (*td*) gene (2,3). Genetic studies (6,21) have demonstrated that this enzyme's expression is essential for the mobility of the *td* intron among the T-even phages, a process that is promoted by duplicative recombination and initiated by an endonucleolytic event (7,8). The I-Tev I as expressed from the IRF is 245 amino acids long. However, one form of the endonuclease isolated previously by us contained only 157 amino acids of the IRF gene plus a 17 amino acid fusion peptide at the amino end (3). This, we have determined, is due to a C to T mutation that converted the 158th codon of the IRF to a stop codon during the course of cloning the IRF into the expression plasmid pET3c. We have successfully purified I-Tev I-157 and I-Tev I, and found them to be equally active in cleaving the *td*ΔI substrate. It is apparent that the last third of the endonuclease (88-residues) is not essential for enzyme activity. I-Tev I-157, like the full-length enzyme molecule, is highly basic due to the large number of basic amino acid residues (9 arginines, 24 lysines and 5 histidines), relative to acidic residues (6 aspartates and 14 glutamates). I-Tev I is even more basic since its ratio of basic to acidic residues is 2.3 relative to 1.9 for I-Tev I-157. Similar to the other known group I intron-encoded endonucleases (e.g. I-Sce I) which are highly basic (22,23), I-Tev I and I-Tev I-157 bind tightly to cationic exchange columns and are eluted only with NaCl concentrations as high as 0.5 M. The very basic nature of the endonuclease causes it to migrate slower than expected on SDS-PAGE gel, with an apparent size of 22,000 in the case of I-Tev I-157 (MW about 20,000) (Fig. 1), and 30,000 in the case of I-Tev I (MW about 28,000) (not shown). It should be stressed that the specific activity for I-Tev I does not take into account inactive and/or partially active enzyme molecules as a result of renaturation.

It was previously shown that an 87-bp region spanning the intron insertion site in the *td*ΔI gene contains the recognition and cleavage signals for the *td* intron endonuclease (4). Through the use of heteroduplex substrates which consisted of M13*td*ΔI ssDNA and site-specific oligonucleotide primers with (Fig. 2) and without T7 DNA polymerase-directed elongation (Fig. 3B), the minimal length of the *td*ΔI gene for cleavage by I-Tev I-157 (and I-Tev I) was determined to be 39 bp in length (region A in Fig. 2), from 27 bp upstream to 12 bp downstream of the intron insertion site (marked by ▽ in Fig. 3A). The cleavage of a 39-bp homoduplex substrate by I-Tev I-157 (Fig. 3C) confirmed that region A alone contains sufficient information for the enzyme to recognize the substrate cleavage site. The requirement of double-stranded structure in the DNA substrate for cleavage by the endonuclease was reaffirmed by the fact that the 39 mer oligonucleotide by itself remained uncleaved (Fig. 3C, lane 5).

There are three major differences between I-Tev I and the I-Sce endonucleases. Firstly, the 39-bp recognition sequence for I-Tev I is more than twice as long as the 18-bp sequence in the cases described for I-Sce I (8) and I-Sce II (23,25). It should be noted that while the cleavage site of I-Sce I is at the intron insertion site (8,22) and that of I-Sce II is only 3 bp downstream (23,25), that of I-Tev I is 24 bp upstream of the intron site (4). The expansive block of nucleotides between the cleavage site and the intron insertion site in the *td*ΔI substrate without doubt contributes to the length of the absolute recognition sequence for I-Tev I.

Secondly, the cleavage sites of both I-Sce endonucleases are centrally situated in their respective recognition sequences. In contrast, I-Tev I cleaves at only 3 bp from the upstream end. This extreme skew of the cleavage site in the recognition sequence may reflect an unusual configuration of one or more recognition (binding) domains relative to the active domain in the I-Tev I endonuclease. In all probability, the enzyme's recognition domains interact mainly with the 24-bp exon1 sequence between the cleavage and the intron site, and with the first 12 bp in exon2 in the DNA substrate.

Thirdly, the group I intron endonucleases found in yeast mitochondria (I-Sce I and II) and in slime mold nucleus (I-Ppo I) exhibit fairly strict specificity, cleaving only their respective intronless homologous alleles at or near the intron homing site. An exception is a form of I-Sce II (pal 4/A) which has been shown to cleave up to two loci in the *E. coli* chromosome (23). The I-Tev I endonuclease, on the other hand, cleaves many intronless TS genes from diverse sources. We have shown in this work that not only is the phage *td*ΔI gene cut by I-Tev I, but also the TS genes from *E. coli*, *L. casei* and human (Fig. 4). In addition, it cleaves the TS gene from two yeast strains though with lower efficiency (data not shown). The mechanism of cleavage of these TS genes derived from very different organisms shows a high degree of similarity, generating in all cases a 2-bp staggered cut containing 3′ hydroxyl overhangs (Fig. 5). Of interest is the fact that the four codons around the cleavage site in the TS genes which are substrates for I-Tev I invariably code for the amino acid sequence -tyr-gln-arg-ser- (Fig. 5). It is more than coincidence that each of the four codons shows variation predominantly in the third nucleotide position. The arg codon exhibits variability in the first nucleotide as well. The increased variability in the third position of codons in the recognition sequence has also been observed for the I-Sce II endonuclease which appears to favor the lateral mobility of introns among gene sequences encoding homologous amino acid sequences (23).

A comparison of the TS gene sequences around the cleavage site shown in Figure 5 has yielded a consensus sequence, 5′-T-A-Py-C-A-Pu-C/A-G-N-T-C-N-3′. To examine the role of this 12-bp substrate sequence in the recognition and cleavage by I-Tev I, *in vitro* oligonucleotide-directed mutagenesis was employed to generate mutant forms of the *td*ΔI sequence for testing as substrates for the endonuclease. Based on the results from mutational studies on six of the twelve positions (Table 1, IV), a tentative sequence for the endonucleolytic cleavage is postulated

as follows: 5'-N-N-N-C-A-N-N-G-N-T/A/C-C-N-3' where the underlined positions have been experimentally tested and the indicated nucleotides verified. From this study, it appears that the degree of sequence degeneracy is rather high. Whereas degeneracy in the first three positions was anticipated based on the determined minimum *td*ΔI substrate length which does not include positions more than 3 bp upstream of the cleavage site (Figs. 2 and 3), degeneracy in the twin positions in the stagger was not. It thus seems that the actual cleavage site possesses little nucleotide specificity. However, downstream of the site at position +4, the T in the *td*ΔI substrate could be substituted by A or C without change in the cleavage efficiency, but when replaced with G, the efficiency was reduced by 50% (Table 1). This indicates that specificity resides mainly in the positions downstream of the cleavage site probably for recognition by I-Tev I. Currently, the remaining six positions in the 12-bp sequence are being examined for their role in recognition and cleavage by the intron endonuclease.

The observed high degree of degeneracy in the substrate sequence for a group I intron endonuclease is not without precedent. The I-Sce II substrate sequence exhibits appreciable degeneracy in that only one (also at position +4) out of 18 nucleotides has been shown to be critical for cleavage (23). In contrast, I-Sce I which shares a common structural dodecapeptide motif (LAGLI-DADG) with I-Sce II, shows a relatively low degree of degeneracy in its substrate sequence (8,22). Such diversity in group I intron encoded endonucleases is emphasized by the fact that the dodecapeptide motif is not found in I-Tev I (2), or in I-Ppo of *Physarum polycephalum* (26). Thus it appears that although the group I introns possess similar intron core structures for splicing, their IRFs when present, code for endonucleases with rather diverse properties. It is reasonable to assume that the IRF-containing introns arose from the invasion of group I introns by endonuclease-encoding ORFs, as suggested in a model for the evolution of mobile introns (5). The evolutionary relationship of the different group I intron endonucleases remains to be determined.

ACKNOWLEDGEMENTS

We thank Angelo Lobo and Santas Seitz for their capable synthesis of the oligonucleotides used in this work and Judy Valentino for her care in the preparation of this manuscript. The work was supported by grants from the National Cancer Institute (CA44355) to F.M. and the National Science Foundation (DMB90-03737) to G.M.

REFERENCES

1. Chu F.K., Maley G.F., Maley F., Belfort M. (1984) *Proc. Natl. Acad. Sci. USA* **80**, 3309–3312.
2. Chu F.K., Maley G.F., West D.K., Belfort M., Maley F. (1986) *Cell* **45**, 157–166.
3. West D.K., Changchien L.-M., Maley G.F., Maley F. (1989) *J. Biol. Chem.* **264**, 10343–10346.
4. Chu F.K., Maley G., Pedersen-Lane J., Wang A.-M., Maley F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 3574–3578.
5. Bell-Pedersen D., Quirk S., Clyman J., Belfort M. (1990) *Nucl. Acids Res.* **18**, 3763–3770.
6. Bell-Pedersen D., Quirk S.M., Aubrey M., Belfort M. (1989) *Gene* **82**, 119–126.
7. Szostak J.W., Orr-Weaver T.L., Rothstein R.J. (1983) *Cell* **33**, 25–35.
8. Colleaux L., d'Auriol L., Galibert F., Dujon B. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6022–6026.
9. Studier F.W., Moffat B.A. (1986) *J. Mol. Biol.* **189**, 113–130.
10. Rosenberg A.H., Lade B.N., Chui D.-S., Lin S.-W., Dunn J.J., Studier F.W. (1987) *Gene* **56**, 125–135.
11. West D.K., Belfort M., Maley G.F., Maley F. (1986) *J. Biol. Chem.* **261**, 13446–13450.
12. Belfort M., Maley G., Pedersen-Lane J., Maley F. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 4914–4918.
13. Pinter K., Davisson V.J., Santi D.V. (1988) *DNA* **7**, 235–241.
14. Davisson V.J., Sirawaraporn W., and Santi D.V. (1989) *J. Biol. Chem.* **264**, 9145–9148.
15. Takeishi K., Kaneda S., Ayusawa D., Shimizu K., Gotoh O., Seno T. (1985) *Nucl. Acids Res.* **13**, 2035–2043.
16. Kenny E., Atkinson T., Hartley B.S. (1985) *Gene* **34**, 335–342.
17. Beardon J.C. Jr. (1978) *Biochim. Biophys. Acta* **533**, 525–529.
18. Zoller M.J., Smith M. (1983) *Meth. in Enzymol.* **100**, 468–500.
19. Taylor J.W., Ott J., Eckstein F. (1985) *Nucl. Acids Res.* **13**, 8765–8785.
20. Laemmli U.K. (1970) *Nature (London)* **227**, 680–685.
21. Quirk S.M., Bell-Pedersen D., Belfort M. (1989) *Cell* **56**, 455–465.
22. Monteilhet C., Perrin A., Thierry A., Colleaux L., Dujon B. (1990) *Nucl. Acids Res.* **18**, 1407–1413.
23. Sargueil B., Hatat D., Delahodde A., Jacq C. (1990) *Nucl. Acids. Res.* **18**, 5659–5665.
24. Wernette C.M., Saldahna R., Perlman P.S., Butow R.A. (1990) *J. Biol. Chem.* **265**, 18976–18982.
25. Wenzlau J.M., Saldahna R.J., Butow R.A., Perlman P.S. (1989) *Cell* **56**, 421–430.
26. Muscarella D.E., Ellison E.L., Ruoff B.M., Vogt V.M. (1990) *Molec. Cell. Biol.* **10**, 3386–3396.

VOLUME 19 NUMBER 24 DECEMBER 25, 1991

Nucleic Acids Research

IRL PRESS
at
OXFORD UNIVERSITY PRESS
Oxford New York Tokyo

ISSN 0305 1048 Coden NARHAD